NHESS: Subseasonal forecasts of heat wave in west African cities
Revisions 2 / review 3

While I have always found the topic of this study interesting and important, and I find this version improved over previous versions, I still have some concerns regarding the methodology and results, and structure of the paper. Some aspects of the methodology and data descriptions are still unclear, and the structure, while improved, is still at times confusing. For example, the skill analysis could do with being split into different parts, each clearly stating the aim, scores chosen, and results, as there seem to be multiple parts to the analysis but it is currently confusing to disentangle them. (e.g. there is one part discussing turning probabilistic forecast to deterministic, and then a description of probabilistic skill scores, with no real clear distinction between the different parts of the analysis and which parts of the results/discussion refer to the different aspects of the study). I have provided some more detailed comments and suggestions below which I hope can be of help to the authors in revising the paper.

Abstract

The abstract is now very long following the additions. While it is great to see more results highlighted in the abstract, it needs to be reduced in length overall. The authors should also consider what information is clear without the context of the full paper. It is not clear what the 15% / 30% and week 5 / week 2 refer to exactly, so this could be omitted as you really need the context of the full results. For example, I would recommend to rewrite the newly added part (end of line 14 onwards) as follows:
"The results suggest that at subseasonal timescales, the forecast models provide a better forecast than climatology, but the hit rate and false alarm rate are sub-optimal and the forecasts may be overestimating the duration of heatwaves, while under-predicting the intensity. Nevertheless, the use of subseasonal forecasts in west African cities can be recommended for prediction of heatwave onset up to two weeks in advance."

Introduction

Line 75- 76: The authors state that Batte showed that the Meteo-France model can predict heatwaves up to one week in advance, in the previous paragraph. Here, they say that this approach can't provide information about the onset and duration of heatwaves. Why is that? It is not clear – if a forecasting system can predict a heatwave up to one week in advance, it implies that the onset could be predicted up to one week in advance. Did Batte et al show specifically that onset and duration are not well predicted? If so, that should be highlighted. Perhaps the authors wish to highlight that while studies have shown that forecast models have skill in predicting heatwaves in the short to medium-range, such information can be well supported by longer-range forecasts which may be able to provide even earlier information about the potential onset of a heatwave event, allowing more time to prepare, and that is what is studied here.

Section 2

The description of the ECMWF forecasts is still a bit confusing to identify which data exactly have been used by the authors, as several different datasets are described. Additionally, the model version used by the authors is not the most recent version of the ECMWF model, which may be confusing in parts because the authors refer to 'is running', and that hindcasts are using 'the most recent version'. This should be re-worded to indicate that the study uses forecasts from the IFS cycle 47r3, which was operational from October 2021 to June 2023.

(for info, the current forecast version is 48r1, released in June 2023. This version consists of a 51-member ensemble forecast with a horizontal resolution of around 9km out to 15 days, and another 51-member ensemble forecast out to 46 days at 36km horizontal resolution. The version will be upgraded again later in 2024)

In the UKMO forecast data description, it states that the initialisation dates are not the same as ECMWF, which does not match with your statement in the ECMWF section that only Thursdays are used to be consistent with the other models. I still do not find a valid reason from the authors for only using the Thursday reforecasts of ECMWF, and not also using the Monday forecasts to increase the sample size. Additionally, the UKMO section is confusing because it states that the hindcasts cover 1993 to 2016, and then starts describing forecasts prior to and since 2017. This needs to be rewritten to be more clear. I would recommend to state the full period of time covered by any UKMO hindcasts, and then break it down into the differences in the forecasts from 2001-2016, and 2017-2021.

Line 203: The authors may wish to state that the land-sea mask of ERA5 has values on a scale of 0 to 1, and what this means, since it is not necessarily standard practice (many masks provide a binary 0 or 1, so it should be clarified what setting a threshold of 0.5 actually means, because it may not be obvious to the reader if they don't work with the ERA5 data and mask themselves)

Heat wave detection – is it a regional characteristic, and documented, that humidity is typically lower during the day? Is the humidity always dry enough that it doesn't have an impact on heat stress during the day? This should be clarified, because it is not the case everywhere – typically, it is important also to consider the humidity during the day in combination with the peak temperatures, in order to truly capture the risk of heat stress. The authors may also want to comment on the impact of both together – it can read as though dry and wet heat waves are two different types of event, but really they are just the daytime and nighttime components of one heatwave event, they are not completely distinct, the authors are just disregarding humidity during the day. The authors may wish to comment on the importance of nighttime temperatures and humidity following extreme daytime temperatures, due to the chance (or lack of) for recovery overnight. What is the reference temperature in the relative humidity formula?

Line 240: "The 90th percentile appears to be a sufficient threshold for monitoring heat waves affecting human health." How did the authors determine this? Can the authors

cite the literature where this is determined? Other studies I have seen use the 95th percentile as a relevant threshold for human health.

Line 251 – 253: is 'hot day' an accurate description, when T2m_min deals with nighttime temperatures? Or is a 'hot day' one where **all three** of the variables are all exceeding their respective 90th percentiles during a 24-hour period?

Lines 256 – 264: I am completely confused by this paragraph, sorry – I don't understand any aspects surrounding the definition of heatwave intensity or how it relates to other aspects. This paragraph needs to be rewritten, but I'm afraid I don't understand well enough to offer any advice on it.

Line 279: "the probability of a member predicting a day as being part of a heatwave" – one ensemble member cannot have a probability attached, as the probability itself comes from the full set of ensemble members; I'm not sure what the authors are trying to say here.

Line 278: why is the threshold of 20% used? The authors state it is inspired by another paper, but a prediction of just a 20% probability of a heatwave could be seen as quite low. It would be important to provide a reason for why 20% is decided to be the optimal threshold and how a different threshold may impact the study.

Lines 283 – 289: what is "the whole season" – what part of the year?

Section 2.4.5: it is confusing to read in the previous section that the authors turn the probabilistic forecast into a deterministic one, because it's too complex to assess probabilistic forecasts, and then immediately the next section introduces probabilistic skill scores.

Results

Line 413-414: "the main differences… are found with ECMWF" – what are the differences between T2m and Tw that are found?

Discussion

Lines 499-503: While the UKMO may place special emphasis on data relevant to the UK and surrounding regions, it does also assimilate a wide range of global and regional data from satellites, ground-based measurements etc. It is unclear how this would influence the surface-ocean interactions. Do the authors mean to discuss differences in ocean coupling between ECMWF and UKMO?

Line 504-505: Is this the correct resolutions for the forecasts that the authors have analysed in this study? The ECMWF extended range is 36km spatial resolution.

Conclusions

Lines 551-554: If these models can be useful only up to two weeks in advance, can the authors comment on why they should be considered for use in heatwave forecasting, if other studies have shown that medium-range forecasts are also accurate up to 2 weeks

ahead for forecasting heatwaves? If choosing between a 2-week medium-range forecast at high resolution, and a lower-resolution subseasonal forecasts up to 2 weeks ahead, why choose the subseasonal forecast?

Figures

Figure 4: the colour scale used in this figure is inappropriate for the data for 2 reasons: (1) the scale shows only negative values from 0 to -14, it is therefore not appropriate to use a diverging colourscale with a centre point at -7, it should use a continuous colour scale, not diverging. (2) the colour scale includes both green and red together, which applying a filter for colour blindness accessibility indicates it is not accessible for those with colour vision deficiencies. Also, are there really not any positive bias values across the entire region, the bias is only negative everywhere? I would find this surprising. I couldn't find any discussion of the results.

Minor comments and typos

Abstract, line 2: remove 'dramatic' (not every heatwave is dramatic, and it is an odd choice of adjective for a heatwave)

Introduction

line 29: change "due to urban heat islands" to "due to the urban heat island effect."

line 30: change "by several heat extreme events" to "by several extreme heat events"

line 31: remove "in the shade" (assuming this is based on an official temperature reading, any such reading should always be in the shade)

lines 45-46: the authors mention that EWS integrate shorter and medium-range forecasts, then say 'this window refers to subseasonal timescale from 2 up to 6 weeks' – but short to medium-range timescales would be defined as up to 2 weeks, and then subseasonal as 2 to 6 weeks. Please rewrite these sentences, for example "Many early warning systems integrate short and medium-range foreacsts of potential weather hazards up to two weeks ahead. The subseasonal forecast range, from 2 to 6 weeks ahead, is also highly relevant for actions aimed at mitigating the consequences of extreme heat"

Line 52: "wet/dry temperatures" is not a typical phrase and is confusing. Recommend changing to "Heat waves are often associated with extreme heat, which can be exacerbated by other factors such as humidity levels."

Line 61: "It refers to…" should be "Heat stress indices are used to combine relevant atmospheric variables (such as temperature, humidity, solar and thermal radiation, wind speed) to indicate the impact of the environment on the human body. Examples include…"

Line 64: Universal Thermal Comfort Index should be "University Thermal Climate Index"

Line 71: "ECMWF extended long-range forecasting system" should be "ECMWF extended-range forecasting system"

Line 74: wet and dry heatwaves are introduced here, but are not defined yet. Consider adding short definitions in brackets e.g. "wet heat waves (those combined with high humidity levels) are more predictable than dry heatwaves (those combined with low humidity levels)."

Line 81: add "subseasonal" before predictability to clarify to the reader the forecast range which you will evaluate

Line 108: “reanalysis data have a high resolution compared to observations” – this is an odd statement, as it is strange to compare a gridded product resolution with the observation network, if sparse. There is also something incorrect here – reanalysis products have too **low** of a resolution to be able to detect the highest temperatures (potentially) at point locations, and the urban heat island effect.